DOCUMENT RESUME

ED 128 440 TM 005 629

TITLE National Council on Measurement in Education Task Force on Measurement of Affective Outcomes. Final Report.

PUB DATE [Apr 76]

NOTE 10p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.

DESCRIPTORS Affective Behavior; *Affective Objectives; *Affective Tests; *Evaluation Methods; Measurement; *Testing Problems; Test Reliability; Test Validity

ABSTRACT

Over forty articles dealing with studies of affective variables were analyzed in this report of the Elementary School Subcommittee of the National Council on Measurement in Education's Task Force on Measurement of Affective Outcomes. The measurement problems were of both a theoretical and practical nature. The greatest practical problem is establishing validity mainly because of problems in conceptualizing the constructs. Another major problem is developing a measurement model. Affective variables are, of their nature, introspective and idiosyncratic. But training has been to deprecate such measurement. The major theoretical problem is the need for different criteria for evaluating these instruments. They should not be judged by the same criteria as cognitive measures because their aims are different. (Author/DW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED128440

Final Report

National Council on Measurement In Education
Task Force On Measurement Of Affective Outcomes

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Elementary School Study Group
Jim Ayrer, Chairman

School District of Philadelphia
Office of Research and Evaluation

Presented at the Annual Meeting of
the National Council on Measurement in Education
April 1976
San Francisco, California

TM005 629

INTRODUCTION

One of the things we as a measurement community do most poorly is the measurement of affective "states." Recognizing this, in 1974 NCME created the Task Force on Measurement of Affective Outcomes. Nancy Cole was asked to chair the group.

It was decided that there were four questions which were of interest and important to answer. They were: (1) which affective variables are Educators most interested in measuring, (2) how are they measuring them, (3) are instruments available, and (4) what are the problems and needs with which NCME could provide assistance. Five study groups were formed to look into these questions at the pre-school, elementary school, secondary school, college, and professional level. This is the report of the elementary school group.

METHOD

To answer these questions required looking into a number of places. The following sources have been used:

A. A search of the ERIC system using as descriptors the terms Affective Behavior, Affective Tests, Evaluation Techniques and Evaluation. The search resulted in 96 citations. Thirty-six seemed to be of value to the aims of this study. These 36 were read to determine:

1. The named variable of interest (e.g., self-concept, attitude toward school)
2. The specific variables (e.g., self-esteem, self-concept as a learner; rules of the school, harmony between races, perception of teacher)
3. How the variables were measured
4. Whether any reliability or validity data were reported
5. Who was tested

6. The objectives of the study

7. The problems encountered in the measurement

Of these, 12 studies had value and were included in this report.

B. A search of the 1975 AERA conference proceedings showed 33 papers which had a major focus on affective variables. Addresses were found for 20 of these authors and copies of their papers were requested. Disappointingly, only five responded. These were analyzed along the dimensions listed under A.

C. There is an Affective Education Special Interest Group in AERA. The leadership of this group put an announcement and request for research findings in their newsletter. Four papers were received and are included in the results.

D. Journal References. A few available journals dated 1970-1975 have been reviewed where the articles have involved measurement of affective variables. Ten pertinent articles were included in the results.

E. Pennsylvania's Educational Quality Assessment (EQA) Battery. This is a set of ten instruments; eight of them, Affective. They are a wide ranging set of instruments and represent a very courageous attempt to get a handle on some things that are extremely difficult to measure (Tolerance, Preparing for a Changing World, Appreciation of Human Accomplishment, and others).

F. Personal Contacts. Philadelphia has a very fine Affective Education Unit that has been functioning since 1968 and has had an evaluation component for all seven years. They have some very good insights into measurement which they shared.

RESULTS

Forty-one studies were analyzed along the seven dimensions previously mentioned. They gave rise to the results listed below.

A. The first question was which affective variables are educators most interested in measuring. That seems to be a function of the researcher's role. Those involved in project evaluation and basic research still seem to be most interested in Self-Concept (variously expressed as Self-Image, Self-Esteem or Self-Concept) and Attitude Toward School; twelve of the former, ten of the latter. Some of the more unusual variables measured were Attitude Towards Pollution and Conservation, Empathy, Impulsivity-Reflectivity, Moral Judgments, Inquiry Skills, Open Mindedness, Attitude towards Smoking, Tolerance, Citizenship, Preparing for a Changing World and Appreciation of Human Accomplishments.

B. The next question was how are they measuring them. Most of the instruments were home grown. Of the 41 instruments used, only five were well known, widely used, quasi-standardized instruments (e.g., Coopersmith, Piers Harris). The techniques used in these instruments were quite varied. There were rating scales, observational checklists, Likert scales, semantic differentials, sentence completion devices, and standardized interviews. Seven of the papers provided information on reliability and validity.

C. Are instruments available? There seem to be only a few instruments of any known quality available, but that apparently doesn't bother anyone. They simply write their own and, in the absence of any information concerning the quality of the instrument, they proceed to tell you the answer to the experimental question.

D. What are the problems and needs with which NCME might provide assistance? The problems, stated and unstated, of the authors fall under

five general headings.

a. Providing Feedback. How do you provide feedback to teachers, parents, students, administrators? How do you present scores to make them meaningful, both in terms of format and score unit? Are scores reliable enough to give feedback on children or should reports be restricted to school or system results? Is a higher score better on Self-Concept than a lower score (some people say if a student is poor at something, his self-concept should reflect it, i.e., a realistic self-concept is better than a good one)? What does a teacher do next when given the results of such a test?

b. Privacy. How do you get the information and still respect the privacy of the individual?

c. The Instrument. How do you make sure the tests are of an appropriate reading level or, if they are read to the students, an appropriate listening vocabulary level? There are no tests of interest or need achievement at the grade 1-3 level? How can you be sure that primary children understand what you are asking (this is not a vocabulary problem as much as it is a problem of their lack of experiences)? How do we avoid social desirability in responding?

d. The Constructs. Some tests measure different things when the same instrument is used at different grade levels. Many tests seem to measure something besides what they're called. How do you test youngsters on things they haven't experienced or where they haven't been able to play the role (important in tests of Citizenship, Interest tests)? There are great problems in conceptualizing the construct and, as a result, problems in defining and measuring the criterion. Therefore, many affective instruments have problems establishing validity. Another problem with validity

is that correlations with other measures of the same construct are frequently low. Where this happens in, for example, self-concept measurement, it is probably because there are so many different dimensions to the construct. Test developers measure different aspects of the same construct and then end up with low correlations.

e. Mode of Measurement. The "scientific" approach stresses the objective and measurable. But affective education emphasizes the idiosyncratic, intuitive, and interpretive. What is the model for measuring such outcomes?

DISCUSSION

Given such a set of problems, it would be possible to go on and on for many pages discussing them. This is not the intent here. Many of the problems are common to all measurement, not just affective measurement. There are, however, some really major problems to which NCME might address itself.

The problems are of both a theoretical and practical nature. The practical problems have been mentioned in the last section. The practical problems of greatest moment seem to be: (1) establishing validity. There are great problems in conceptualizing the constructs. The result of that is great problems in writing items that measure the construct and the dimensions of the construct that we wish to measure. It is also hard, as was mentioned before, to identify or develop criterion measures that get at the construct. An excellent set of examples is Pennsylvania's EQA battery. This is a fine attempt to measure a series of variables that most school people would agree represent important goals of any school system: Tolerance, Citizenship, Appreciation of Human Accomplishments, Preparing for a Changing World. In most cases, the test developers

attempted to use a peer or teacher rating as a criterion, and, in most cases, the correlations were less than .25.

One might think that the problem is the unusual nature of the variables, the fact that there is not much experience in attempting to deal with them, and that is in part true. But we haven't had much more success with variables such as Self-Concept and Attitude Towards School. We just don't seem to have the language or terminology to describe and distinguish goals and objectives in the affective area. What do people really have in mind when they talk about improving self-concept or attitude towards school? These are two of the most frequently measured affective variables, but there are many different kinds of self-concept and different facets of attitude towards school. We in the measurement community should be developing a taxonomy to make sure we are all talking about the same thing.

(2) A second major practical problem is developing the model for measuring affective variables. Cronbach talks about the Fidelity-Bandwidth Dilemma. The more narrow the bandwidth, the greater the fidelity with which we can measure. But the kinds of things we want to measure in the affective area are of a very broad bandwidth. Much of what we are trying to measure is introspective, and idiosyncratic. We have been trained to look down our noses at this kind of measurement but, if we believe in the importance of this kind of education, we have to find a way to make it respectable. It's strange, too, because that's where Psychology started 100 years ago. Once we've made it respectable again (or concurrently), we then have to develop the rules for constructing that kind of instrument and evaluating it.

This leads us to the theoretical problem. How are these instruments to be evaluated. What are the criteria against which the psychometric

qualities should be judged? The criteria for evaluating instruments in our field were developed from the perspective of using cognitive instruments to make decisions about individuals. But the needs here are to use affective instruments to make primarily group decisions. It seems, therefore, that we need a rethinking of these criteria. This is especially true as it refers to reliability and validity. Reliability for individuals is not as important as reliability for groups while the key to the instrument's value is still it's validity.

Two final things must be done. First is the dissemination of all the information that has been developed as a result of the previous steps. Second is developing methods of providing feedback to school personnel so that the results can be used.

This latter has some subtle problems because we tend to make certain value judgments about things like self-concept, attitude towards school, and the like. Is a high score on self-concept the aim for all students? As was mentioned previously, some people believe that a realistic, low self-concept is better than an unrealistic, high one. If some students don't like school (a realistic, sensible attitude for some students who find school frustrating and unpleasant), are they wrong? Has a project been a failure if it doesn't produce a high attitude towards school among high school students who are functioning at a third grade level in reading and math? If a test shows a student to be intolerant, a bigot, are teachers required to change that? There is something wrong with outsiders making that decision for a student.

NCME's ROLE

NCME's role must now become a more active one. Groups of interested researchers and test developers should be constituted to attack each of

these problems. Given the impecunious nature of NCME, the rewards for these groups cannot be financial, so they must be psychic. This can be accomplished by guaranteeing space in JEM, devoting an issue of Measurement News to the measurement of affective variables, providing a showcase at the annual meeting, and the like. This will give the committees an outlet for their work and provide for dissemination at the same time.